KEKER VAN NEST & PETERS   **LATHAM&WATKINS**LLP   **MORRISON FOERSTER**

November 21, 2024

Hon. Ona T. Wang
United States Magistrate Judge
Southern District of New York
New York, New York 10007

> Re: *The New York Times Company v. Microsoft Corporation, et al.*,
>   Case No.: 23-cv-11195-SHS-OTW: Motion to Compel concerning Various RFPs

Dear Magistrate Judge Wang:

The Times presents a smorgasbord of unripe issues to this Court. OpenAI has either resolved or agreed to produce documents in response to most of these RFPs. The parties are not at an impasse on the rest. The Court should deny The Times's motion ("Mot.").

RFP 2 (government proceedings). OpenAI has already agreed to produce its final served responses for the requested interrogatories before the Federal Trade Commission. Without conceding that this request seeks relevant or proportional documents, there are no other U.S. government proceedings where OpenAI has produced documents "that concern or relate to the allegations in the Complaint." (RFP 2). OpenAI considers this RFP resolved.

RFP 17 (paywalls) & 24 (value of journalism): The parties agree these disputes are resolved.

RFP 49 (AI models "containing" news): RFP 49 seeks documents regarding OpenAI's alleged knowledge of its AI Models "contain[ing] copyrighted content, including Times Content." The Times later clarified in a meet-and-confer that it seeks documents about whether journalism content was included in the Training Datasets for the at-issue models. OpenAI has agreed to produce documents concerning each of the News Plaintiffs and the presence (or lack thereof) of their copyrighted works in the training data for OpenAI's relevant models. OpenAI also agreed, including in separate responses to RFPs 74-80, to produce documents about how and why OpenAI curates and uses different types of data to train OpenAI's relevant models. That includes OpenAI's use of third-party textual news data. Thus, contrary to the News Plaintiffs' speculation (ECF 323 at 2), OpenAI is searching for (and *not* withholding) documents related to third-party written news publications.

The only dispute – which The Times acknowledges was not ripe "as of this filing" – relates to The Times's expansive definition of "journalism." If The Times were to define journalism as the "activity of writing … for newspapers, magazines, news websites … or any other [written] publication," then OpenAI would agree to that scope. The issue is that RFP 49 requests documents about all "copyrighted content," which The Times has defined to *also* include documents about "creating content for … mobile applications, television, podcasts, or any other publication." OpenAI objects to that practically unconstrained breadth because, among other reasons, audiovisual media is not at issue in the News Plaintiffs' complaints. *See Tripathy v. Schneider*, No. 21-cv-6339, 2024 WL 4528763, at *8 (W.D.N.Y. Oct. 18, 2024) ("While discovery is broad and permissive, it must relate to the claims of the operative complaint."). The Court should limit the scope to written news and deny The Times' motion.

RFP 86 (creation of synthetic data): RFP 86 asks for "All Documents concerning use of Generative AI Models to create Synthetic Datasets." The Times's letter-brief, however, does not support production of "all documents" about "synthetic datasets," or even all journalism-related documents (an ask it never made during conferrals). Instead, The Times narrows the focus to The Times's content. *See* Mot. at 2-3. OpenAI agrees that tying this request to Plaintiffs' copyrighted content is an appropriate scope. *Comprehensive Habilitation Servs., Inc. v. Com. Funding Corp.*, 240 F.R.D. 78, 83 (S.D.N.Y. 2006) (discovery must "bear[] on, or that reasonably could lead to other matter that could bear on, any issue that is or may be in the case"). Accordingly, OpenAI has agreed to produce documents sufficient to show the use of OpenAI models, if any, trained on Times Content to create synthetic datasets. The Times now takes issue with the "sufficient to show" limitation. But that language is meant to reflect Plaintiffs' narrowing and the parties' compromise, and thus prevent an unbounded inquiry into the scope of RFP 86 as written – *i.e.*, the development of every piece of synthetic data ever created on any subject and for any purpose.

RFP 87 (contemplated uses of synthetic data): RFP 87 asks for "All Documents concerning [OpenAI's] use or contemplated use of Synthetic Datasets, including to train Your Generative AI Models." Despite this broad written scope, The Times's letter-brief focuses its request on documents related to the market or substitution value (either now or in the future) of OpenAI's synthetic data, as compared to "authentic journalism content" from The Times or other textual news sources. Mot. at 3. This focus is fine: OpenAI is willing to produce documents sufficient to show value assessments for synthetic data as compared to written news. OpenAI's willingness to produce the foregoing stands in stark contrast to The Times's continual refusal to produce market-related discovery. Indeed, OpenAI currently has a pending motion addressing The Times's refusal to produce, *inter alia*, documents related to the reasons for changes in revenue and subscriptions, and the alleged markets for licensing the Times's works. *See* ECF 310.

To be clear, OpenAI's "sufficient to show" language is not included to withhold market or value documents related to written news. Rather, the intent is to focus Plaintiffs' unbounded request for "all documents concerning [OpenAI's] use or contemplated use of synthetic datasets" to what The Times agrees is the narrower scope of this case. OpenAI uses—and contemplates using—synthetic data for purposes that do not relate to textual news and are not substitutive of anything. *See, e.g.*, Ex. 1 at 4 (OpenAI white paper describing synthetic data use for "data augmentation" and "when there is little to no initial data [available]").

As this Court has recognized, discovery requests like RFP 87 "would be proportional if simply framed as 'Documents sufficient to show,' but inclusion of the word 'All' at the beginning is at odds with the notion of providing only a limited set of materials sufficient to show the particular fact." *Hedgeye Risk Mgmt., LLC v. Dale*, No. 21-cv-3687, 2023 WL 4353076, at *2 (S.D.N.Y. July 5, 2023). OpenAI therefore requests the Court accept OpenAI's "sufficient to show" limitation aimed at the core of what The Times seeks, and then deny The Times's motion.

RFP 95 (Data Working Group): RFP 95 seeks "documents and communications concerning" OpenAI's "data working group" ("DWG") with Microsoft, regardless of the topics of those communications or their relevance to this case. This is overbroad. But more importantly, OpenAI should not be compelled to collect and review documents authored by lawyers, on topics relating to legal issues, that were discussed among parties with a common legal interest.[1]

Any responsive, relevant documents from the DWG are almost certain to be privileged. All of OpenAI's members of the DWG were lawyers acting in a legal capacity. *See* Pantuliano Decl. at ¶ 6. And to OpenAI's knowledge, Microsoft's members were members of the Microsoft legal team. *See id.* at ¶ 8.[2] The DWG was formed after OpenAI and Microsoft had already been sued in a putative class action lawsuit relating to data acquisition practices. *See id.* at ¶ 5b; *J. Doe 1 et al. v. GitHub, Inc. et al.*, No. 4:22-cv-06823-JST (N.D. Cal.).[3] The DWG's purpose is to discuss OpenAI's common legal interest with Microsoft. *See id.* at ¶¶ 4-5. This is a paradigmatic common legal interest, both because of ongoing litigation and because of the background legal environment implicated by OpenAI's and Microsoft's work. *See, e.g.*, *United States v. Schwimmer*, 892 F.2d 237, 243 (2d Cir. 1989) ("The joint defense privilege, more properly identified as the 'common interest rule,' has been described as 'an extension of the attorney client privilege.' It serves to protect the confidentiality of communications passing from one party to the attorney for another party where a joint defense effort or strategy has been decided upon and undertaken by the parties and their respective counsel.") (citation omitted); *see also United States v. BDO Seidman LLP*, 492 F.3d 806, 816 (7th Cir. 2007) (finding "joint venturers[] shared a common legal interest 'in ensuring compliance with the new regulation issued by the IRS,' and in making sure that they could defend their product against potential IRS enforcement actions") (citation omitted).

Rather than seriously contesting the applicability of privilege and the common interest doctrine, the Times's request appears to be a backdoor attempt to add more custodians. None of the current or additional requested OpenAI custodians are members of the DWG. *See* Pantuliano Decl. at ¶ 6. If current custodians communicated with DWG members about topics otherwise responsive to issues in this case, those communications will already be either logged (if privileged) or produced (if not). Responding to RFP 95 would impose on OpenAI the burden of collecting, searching, and reviewing attorney documents, only to result in a lengthy privilege log. Such an effort is not proportional to the needs of the case.

---

[1] This same issue is the subject of simultaneous letter-briefing in the related *Authors Guild* matter. *See Authors Guild v. OpenAI*, No 1:23-cv-08292-SHS-OTW, ECF No. 267. OpenAI incorporates by reference its opposition there, including its alternative request to submit full briefing regarding privilege and its common legal interest with Microsoft.

[2] While the Times claims that a cited document "proves that non-lawyers were involved in creating the group and establishing its mission," that assertion is based on an OpenAI employee's incorrect reference. *See* Pantuliano Decl. at ¶ 10.

[3] "A formal written common interest agreement is not necessary." *See, e.g.*, *United States v. Zhu*, 77 F. Supp. 3d 327, 330 (S.D.N.Y. 2014). Regardless, OpenAI and Microsoft memorialized their common interest in a formal written agreement a short time after the first meeting. *See* Pantuliano Decl. at ¶ 5a.

Sincerely,

KEKER, VAN NEST & PETERS
LLP

LATHAM & WATKINS LLP

MORRISON FOERSTER LLP

/s/Andrew Bruns
Andrew Bruns

/s/Elana Nightingale Dawson
Elana Nightingale Dawson

/s/Carolyn Homer
Carolyn Homer